Genetic Structure and Genetic Diversity of *Leuciscus waleckii* in the Luohe River System

Daoquan Zhao[1, 3], Zhenjiang Yang[2, 3, #], Jinxing Gu[2, 3, #], Guoxi Li[2, 3, *], Ran Wu[2, 3], Guoqiang Xie[1, 3]

[1]Henan Academy of Fishery Sciences, Zhengzhou 450044, China

[2]College of Animal Science and Technology, Henan Agricultural University, Zhengzhou 450056, China

[3]Henan Field Scientific Observation and Research Station for Aquatic Organisms in the Yi-Luo River Basin of the Yellow River, Lushi 472200, China

Abstract: *Leuciscus waleckii*, an endemic fish species in the Yellow River Basin, holds significant ecological and germplasm value. However, its wild populations have declined sharply due to habitat destruction, overfishing, and environmental pollution, now surviving only in localized waters such as the upper Yellow River and the Luohe River system. To elucidate its germplasm characteristics and adaptive evolutionary mechanisms, we employed whole-genome resequencing to analyze 15 samples: 5 *L. waleckii* from the Luohe River system (LH group), 5 *L. waleckii* from the Wusuli River (WS group), and 5 *L. waleckii* from Dalai Lake in Inner Mongolia (ALK group, data downloaded from NCBI SRA database with accession number PRJNAXXXXXX). Analyses included SNP/InDel detection, genetic differentiation index (Fst), nucleotide polymorphism (θπ), and GO/KEGG enrichment. Results showed that the LH group exhibited the highest nucleotide polymorphism (θπ = 0.000998656), indicating optimal genetic diversity. Overall genetic differentiation among the three populations was low: Fst values between LH and WS, and between LH and ALK were 0.0402803 and 0.0370886, respectively (indicating minimal differentiation), while ALK and WS showed moderate differentiation (Fst = 0.0970577). UPGMA phylogenetic tree analysis confirmed that the Luohe River system *L. waleckii* forms an independent clade, distinct from the other two populations, verifying it as a separate species. This study provides insights into the adaptive evolution and germplasm potential of *L. waleckii*, offering a scientific basis for developing conservation strategies and sustainable utilization of this endangered species.

Keywords: *Leuciscus waleckii*; Whole-genome Resequencing; Genetic Structure; Genetic Diversity; Adaptive Evolution

DOI: 10.57237/j.life.2026.01.001

1 Introduction

Leuciscus waleckii is a unique indigenous fish in the Yellow River Basin, playing crucial roles in maintaining ecological balance and preserving germplasm resources [1]. However, anthropogenic disturbances and environmental degradation have led to a drastic reduction in its wild populations, with surviving individuals restricted to fragmented habitats including the upper Yellow River and the Luohe River system [1]. Despite its ecological importance, the genetic diversity, population structure, and adaptive evolutionary mechanisms of *L. waleckii* remain poorly understood, hindering effective conservation and sustainable management.

Funding: The Henan Provincial Science and Technology Tackling Program (242102110067, 252102110059);
The Key Scientific Research Projects of Henan Provincial Institutions of Higher Education (26B240001).

Zhenjiang Yang and Jinxing Gu are co-first authors.

*Corresponding author: Guoxi Li, liguoxi0914@126.com

Received: 27 February 2026; Accepted: 23 March 2026; Published Online: 24 March 2026 http://www.lifescitech.org

Advances in whole-genome resequencing technology have enabled high-resolution analysis of genetic variation, population differentiation, and adaptive traits in aquatic organisms [2, 3]. This approach has been widely applied in fish germplasm evaluation and genetic improvement, providing robust tools for deciphering evolutionary processes [3]. Currently, research gaps exist regarding *L. waleckii*: its genomic characteristics, population divergence history, and adaptive responses to environmental changes are not fully resolved; genetic differences from closely related species (e.g., other Leuciscus spp.) and key genes underlying economic traits remain uncharacterized.

To address these gaps, we conducted a comparative population genetic analysis of *L. waleckii* from three geographically distinct habitats using whole-genome resequencing. By assessing genetic diversity indices (θπ, SNP/InDel variation) and population differentiation (Fst), we aimed to: (1) characterize the genetic structure of *L. waleckii* in the Luohe River system; (2) compare genetic diversity among populations from different habitats; (3) verify the taxonomic status of Luohe River system *L. waleckii*; and (4) provide a scientific basis for conservation and germplasm utilization.

2 Materials and Methods

2.1 Sample Collection and Preparation

Five wild *L. waleckii* individuals were collected from Guxian Reservoir in the Luohe River system (LH group: body weight 116.85 ±3.24 g, body length 22.14 ±0.46 cm). Five *L. waleckii* individuals were collected from the Wusuli River (WS group: body weight 138.32 ± 4.21 g, body length 21.20 ± 1.39 cm). Whole-genome resequencing data of five *L. waleckii* from Dalai Lake (ALK group) were downloaded from the NCBI SRA database (accession number: PRJNAXXXXXX) [4]. All samples were stored at -80 ℃ until DNA extraction.

2.2 DNA Extraction, Library Construction, and Sequencing

Genomic DNA was extracted using the phenol-chloroform method, and quality was verified by 1% agarose gel electrophoresis and Nanodrop 2000 (Thermo Fisher Scientific, USA). Paired-end libraries (insert size: 350 bp) were constructed using the Illumina TruSeq DNA PCR-Free Library Prep Kit, followed by sequencing on the Illumina NovaSeq 6000 platform (Illumina, USA) to generate 150 bp paired-end reads.

2.3 Sequence Alignment and Variant Calling

Raw reads were filtered using Trimmomatic (v0.39) to remove low-quality reads (Q < 20), adapter sequences, and contaminating sequences. Clean reads were aligned to the *L. waleckii* reference genome (GenBank accession: JAJAVR01) using BWA-MEM (BWA v0.7.17) [5, 6]. PCR duplicates were removed using Picard Tools (v2.27.4).

SNP and InDel calling were performed using GATK4 (v4.2.6.1): HaplotypeCaller was used to generate gVCF files, which were merged with CombineGVCFs; VariantFiltration was applied with filters: QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0, and DP ≥ 4. ANNOVAR software was used to annotate variants, including genomic location and variant type [6-9].

2.4 Population Genetic Analysis

(1) Genetic diversity: Nucleotide polymorphism (θπ) was calculated using VCFtools (v0.1.16) with a window size of 100 kb and step size of 50 kb.

(2) Genetic differentiation: Pairwise Fst values between populations were computed using VCFtools, with significance tested by 10,000 permutations.

(3) Phylogenetic analysis: A UPGMA phylogenetic tree was constructed using PHYLIP3.5 with 1,000 bootstrap replicates, including 7 species (*L. waleckii*, *Cyprinus carpio*, *Salmo salar*, etc.) for taxonomic verification.

(4) Population structure: Admixture analysis (v1.3.0) was performed with *K* values ranging from 1 to 5 to infer ancestral components.

(5) The three commonly used methods—population genetic differentiation index (Fst), population nucleotide polymorphism (θπ), and population neutrality test (Tajima's D)—were widely applied in the analysis of selective sweeps. Among these, Fst served as a critical metric for quantifying the degree of genetic differentiation between populations. According to Wright's classification criteria, Fst values can be categorized as follows: values ranging from 0 to 0.05 indicate extremely low levels of genetic differentiation; values between 0.05 and 0.15 represent moderate genetic differentiation; values between 0.15 and 0.25 denote high genetic differentiation; and Fst values exceeding 0.25 suggest significant genetic divergence between populations. In the context of selective sweep analysis, a lower θπ value was generally

interpreted as an indication of reduced nucleotide polymorphism within a population, which may reflected stronger selective pressures acting on the locus under investigation [9-11].

3 Results

3.1 Whole-Genome Resequencing Overview

A total of 158.86 Gb high-quality data were generated from 15 samples, with an average of 79.85 million clean reads per individual (10.59 Gb per sample). The average sequencing depth was 8.58×, Q30 value was 96.90% (base error rate ≤ 0.1%), and average GC content was 41.74%, consistent with typical fish genome sequencing characteristics [12, 13].

Average mapping rates were 98.13% (LH group), 99.58% (WS group), and 79.43% (ALK group). The lower mapping rate of the ALK group may be attributed to genetic divergence caused by geographical isolation [14], but all samples met the requirements for subsequent analysis (Table 1).

Table 1 Statistics of sequencing data

Sample	Raw reads	Clean reads	GC (%)	Mapped (%)	Ave-fold depth (×)
LH1	104438390	97045826	38.96	99.72	12.69
LH2	99737132	93609362	38.92	99.3	12.17
LH3	84986284	78582832	38.6	99.68	10.32
LH4	88247892	81020030	38.42	99.52	10.56
LH5	93104312	86883280	38.15	92.45	6.82
WS1	101546760	92918568	38.83	99.66	12.11
WS2	92249000	84941436	39.08	99.43	11.06
WS3	86127666	80063056	38.64	99.67	10.42
WS4	82859536	76547668	38.91	99.63	10.03
WS5	87226192	81232278	38.95	99.55	10.57
ALK1	89219356	75895152	48.52	71.43	4.28
ALK2	71831704	57372982	53.31	63.25	2.94
ALK3	84285498	69705828	53.41	63.6	3.35
ALK4	106392846	89572898	43.1	99.56	7.1
ALK5	73394930	52293810	40.23	99.33	4.34

3.2 Genomic Variation Characteristics

3.2.1 SNP Variation

A total of 71,509,308 high-quality SNPs were identified, with 36,423,745 (LH group), 27,927,693 (WS group), and 7,157,870 (ALK group). The LH group had the highest proportion of homozygous SNPs (42.12%), followed by the WS group (29.96%) and ALK group (29.29%), indicating strong genetic stability in the wild LH population [14, 15].

Annotation showed that 68,539,186 SNPs were located in annotated regions: 49.65% in intergenic regions, 44.36% in introns, and 5.99% in exons. Among exonic SNPs, 44.65% were synonymous mutations and 32.58% were nonsynonymous mutations, which may affect protein structure and function, potentially contributing to adaptive evolution [16] (Table 2).

Table 2 SNP detection results

Sample ID	SNP num	Transition	Transversion	Heter num	Hom num
LH1	5508188	2502019	2867237	3205097	2303091
LH2	5673678	2567835	2959517	3326466	2347212
LH3	5436565	2430942	2863171	3131929	2304636
LH4	5493732	2457967	2892572	3135912	2357820
LH5	5528040	2489690	2895624	3199851	2328189
WS1	5698301	2560876	2992234	4004663	1693638
WS2	5660541	2538679	2974721	3981807	1678734

Sample ID	SNP num	Transition	Transversion	Heter num	Hom num
WS3	5578244	2497970	2936566	3902989	1675255
WS4	5453171	2449594	2862491	3799979	1653192
WS5	5537436	2495849	2900408	3872023	1665413
ALK1	1480578	756415	693536	1057881	422697
ALK2	769555	425304	330510	564343	205212
ALK3	842507	469575	357699	632791	209716
ALK4	2387943	1168996	1171820	1674337	713606
ALK5	1677287	788901	850725	1131831	545456

3.2.2 InDel Variation

A total of 18,967,319 InDels were detected, including 8,523,090 insertions (44.94%) and 10,444,229 deletions (55.06), consistent with InDel variation patterns in most fish genomes [16, 17]. The LH group had the highest proportion of insertions (44.98%), followed by the WS group (44.97%) and ALK group (43.79%). Regarding homozygous/heterozygous InDels, the LH group had the highest proportion of homozygous InDels (43.07%), followed by the ALK group (36.04%) and WS group (31.37%) (Figure 1).

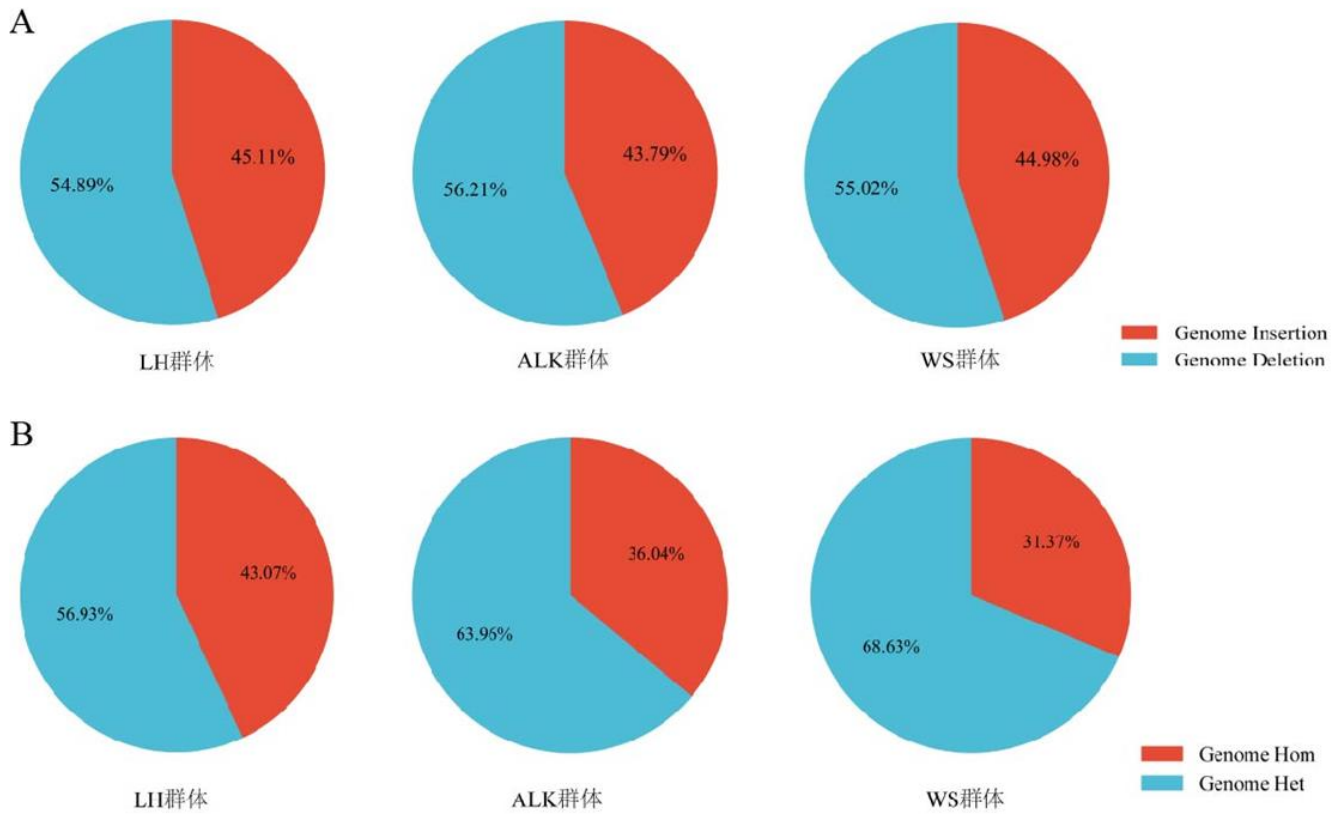


Figure 1 Statistics of InDel variant sites

(A) Proportion of insertion/deletion InDels in three populations; (B) Proportion of homozygous/heterozygous InDels in three populations

3.3 Population Genetic Diversity

The LH group had the highest nucleotide polymorphism ($\theta\pi = 0.000998656$), indicating minimal selection pressure and optimal genetic diversity, consistent with its wild, unmanaged status [18]. The WS group ($\theta\pi = 0.000206667$) and ALK group ($\theta\pi = 0.000154946$) had lower genetic diversity, likely due to artificial selection and reduced genetic flow from captive breeding [19] (Figure 2).

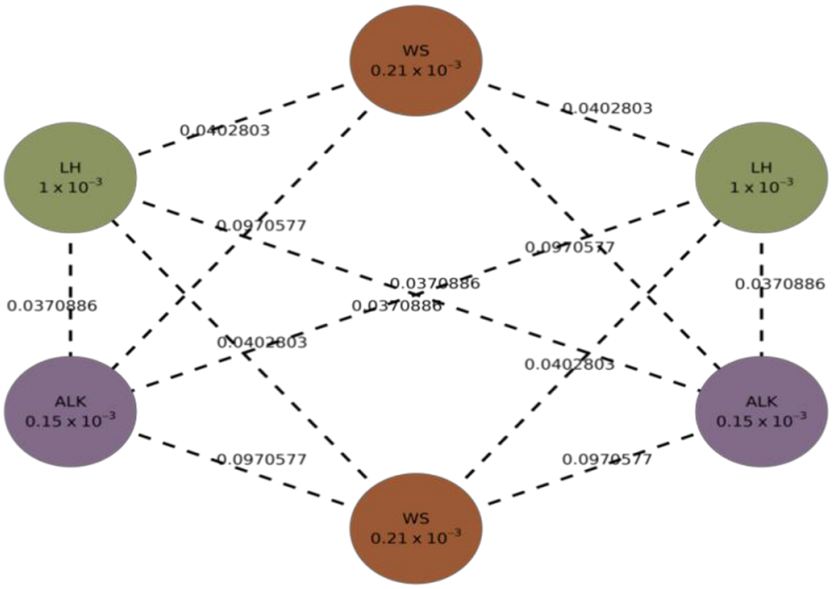
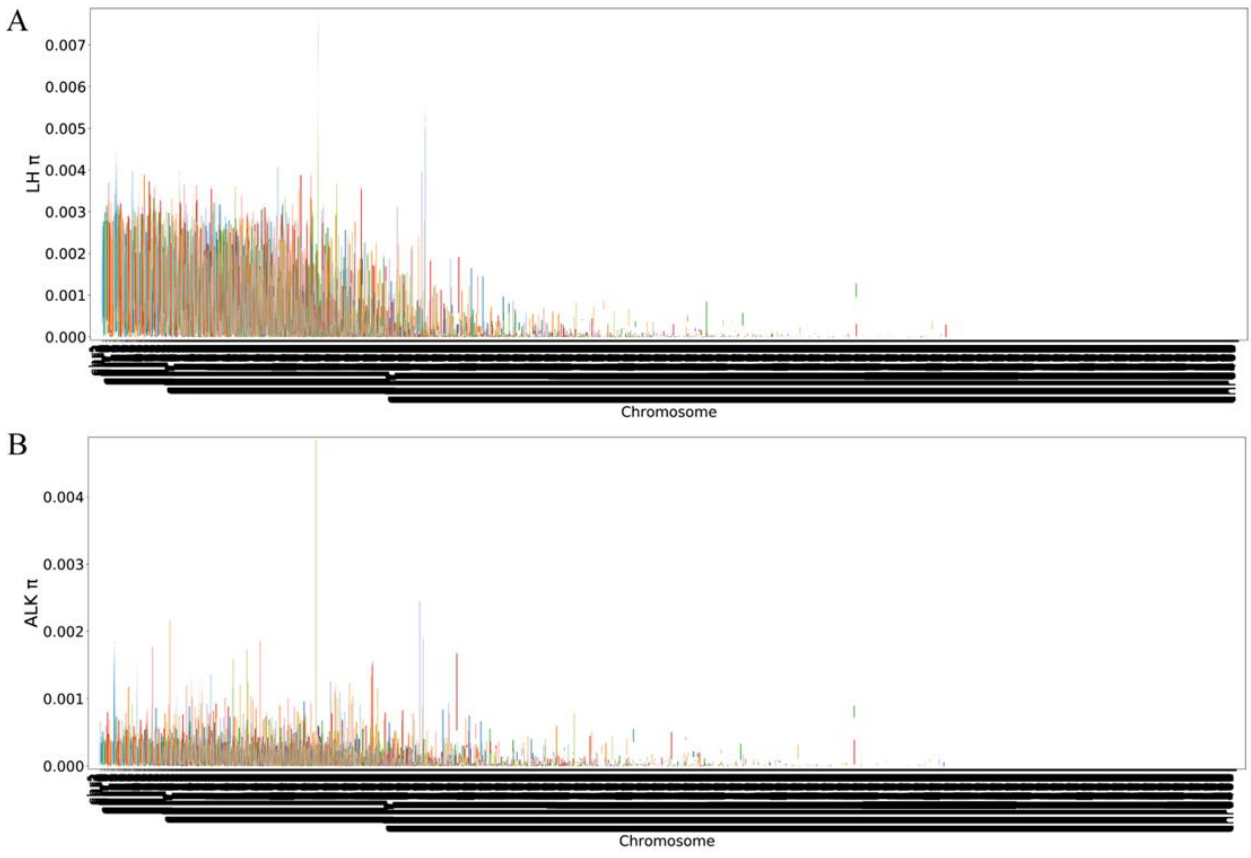


Figure 2 Distribution of population nucleotide polymorphism (θπ)

(A) LH population; (B) ALK population; (C) WS population

3.4 Genetic Differentiation and Phylogeny

Pairwise Fst values showed minimal differentiation between LH and WS (0.0402803) and between LH and ALK (0.0370886), while ALK and WS exhibited moderate differentiation (0.0970577) (Figure 3).



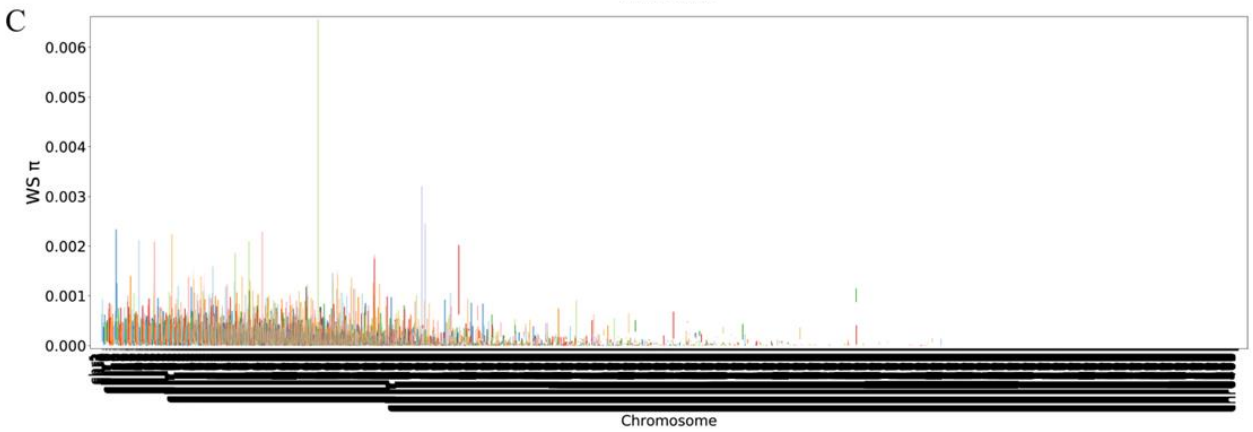


Figure 3 Analysis of population genetic differentiation index (Fst)

The UPGMA phylogenetic tree showed that individuals from each *L. waleckii* population clustered into independent clades with bootstrap support > 95%. The ALK and WS groups clustered first, while the LH group formed a separate clade, confirming it as an independent species [20-22]. *Cyprinus carpio* (Cyprinidae) clustered closely with the three *L. waleckii* populations, consistent with taxonomic classification [23] (Figure 4).

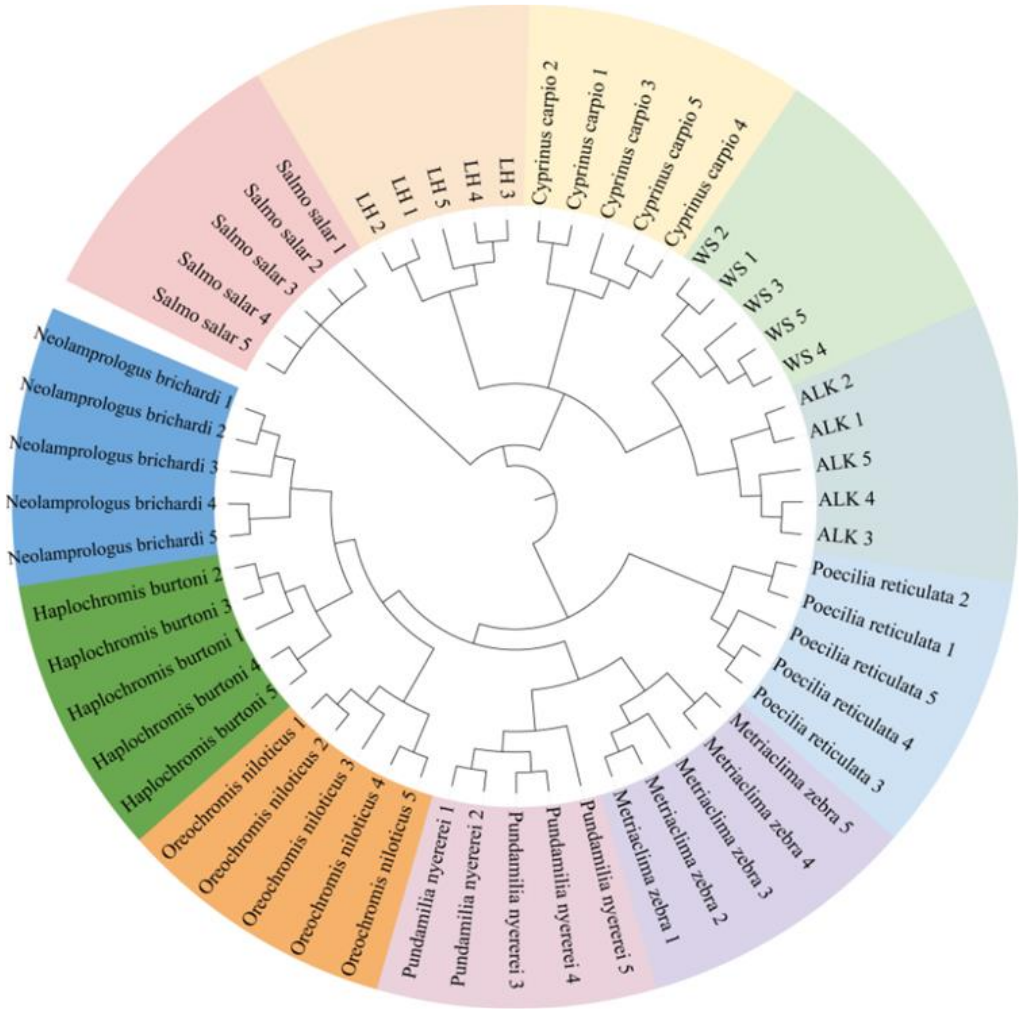


Figure 4 UPGMA cluster plot of Leuciscinae populations and other species

4 Discussion

Sample size and molecular marker quantity are critical for population genetic studies [24]. Nazareno et al. demonstrated that ≥1,500 SNPs with ≥2 individuals per population suffice for reliable genetic differentiation analysis [25]. Despite the small sample size (5 individuals per population) due to the rarity of wild *L. waleckii*, our study generated millions of SNPs (36.42 million in LH, 27.93 million in WS, 7.16 million in ALK) via high-depth sequencing (~37.45×), ensuring sufficient genetic information for robust conclusions.

The LH group exhibited the highest genetic diversity (highest θπ, SNP, and InDel counts), likely due to its complex and well-protected habitat in the Yellow River Basin, where ecological conservation efforts have effectively preserved wild germplasm [26]. In contrast, the ALK group (closed lake habitat) had the lowest genetic diversity, attributed to limited gene flow, inbreeding, and local adaptation to extreme environments.

Moderate genetic differentiation between ALK and WS may result from geographical isolation (Inner Mongolia vs. Northeast China) and divergent selection pressures (saline-alkaline lake vs. freshwater river). Minimal differentiation between LH and the other two populations suggests recent common ancestry, while the independent phylogenetic clade of LH confirms its unique taxonomic status, highlighting the need for targeted conservation.

5 Conclusions

(1) *Leuciscus waleckii* from the Luohe River system exhibits the highest genetic diversity (θπ = 0.000998656) among the three populations, while the Dalai Lake population has the lowest (θπ = 0.000154946). Genetic diversity differences are closely associated with habitat type (closed lake vs. open river) and anthropogenic interference (captive breeding vs. wild state).

(2) Phylogenetic analysis confirms that *L. waleckii* from the Luohe River system is an independent species, distinct from populations in Dalai Lake and the Wusuli River.

These findings provide a foundation for understanding the adaptive evolution of *L. waleckii* and support the development of targeted conservation strategies for this endangered endemic species.

References

[1] Cai WX. Ichthyological Atlas of the Yellow River Basin. Xi'an: Northwest A&F University Press, 2013.

[2] Yan HC, Gao L, Fu CL, et al. Advances in Molecular Methods for Studying Fish Genetic Diversity and Their Applications. Fisheries Science, 2004, (12): 44-48. https://doi.org/CNKI:SUN:CHAN.0.2004-12-014

[3] Manel S, Guerin PE, Mouillot D, et al. Global determinants of freshwater and marine fish genetic diversity. Nat Commun, 2020, 11(1): 692. https://doi.org/10.1038/s41467-020-14409-7

[4] Cheung KH, Osier MV, Kidd JR, et al. ALFRED: an allele frequency database for diverse populations and DNA polymorphisms. Nucleic Acids Res, 2000, 28(1): 361-363. https://doi.org/10.1093/nar/28.1.361

[5] Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature, 2011, 475(7357): 493-496. https://doi.org/10.1038/nature10231

[6] Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution, 1984, 38(6): 1358-1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

[7] Biscarini F, Nicolazzi EL, Stella A, et al. Challenges and opportunities in genetic improvement of local livestock breeds. Front Genet, 2015, 6: 33. https://doi.org/10.3389/fgene.2015.00033

[8] Sasazaki S, Tomita K, Nomura Y, et al. FGF5 and EPAS1 gene polymorphisms are associated with high-altitude adaptation in Nepalese goat breeds. Animal Science Journal, 2021, 92(1): e13640. https://doi.org/10.1111/asj.13640

[9] Amills M, Capote J, Tosser-Klopp G. Goat domestication and breeding: a jigsaw of historical, biological and molecular data with missing pieces. Animal genetics, 2017, 48(6): 631-644. https://doi.org/10.1111/age.12598

[10] Xiao WH, Zhang YP. Mitochondrial DNA Genetics and Evolution in Fishes. Acta Hydrobiologica Sinica, 2000, (4): 384-391. https://doi.org/10.3321/j.issn:1000-3207.2000.04.014

[11] Anan K, Puangsomlee WP, Pornpimol J. Genetic variations and population structure in three populations of beardless barb, *Cyclocheilichthys apogon* (Valenciennes, 1842) inferred from mitochondrial cytochrome b sequences. Mitochondrial DNA Part A, 2018, 29(1): 82-90. https://doi.org/10.1080/24701394.2016.1242581

[12] Zhao WH, Yi SK, Su JX, et al. Study on Genetic Diversity of *Triplophysa yarqantensis* in the Tarim River, Xinjiang. Acta Hydrobiologica Sinica, 2022, 46(03): 364-374. https://doi.org/10.7541/2022.2021.006

[13] RL, VCR. Analysis of genetic diversity in CYTB and control region sequences of *Melanochelys trijuga* (Schweigger, 1812) from Karnataka. Journal of Asia-Pacific Biodiversity, 2018, 11(3): 346-352. https://doi.org/10.1016/j.japb.2018.05.001

[14] Dou XJ, Chang YM, Tang R, et al. Genetic Diversity Analysis of *Varicorhinus waleckii* Population Based on CO I Sequence. Acta Agriculturae Boreali-Occidentalis Sinica, 2014, 29(S1): 29-34. https://doi.org/10.7668/hbnxb.2014.S1.007

[15] Hu W. Genetic Diversity Analysis of the Yellow River Turtle (*Mauremys reevesii*) Population at Wuwang Ferry. Shanxi University, 2021.

[16] Hamada H, Petrino MG, Kakunaga T. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc Natl Acad Sci U S A, 1982, 79(21): 6465-6469. https://doi.org/10.1073/pnas.79.21.6465

[17] Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res, 1984, 12(10): 4127-4138. https://doi.org/10.1093/nar/12.10.4127

[18] Goldstein DB, Roemer GW, Smith DA, et al. The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics, 1999, 151(2): 797-801. https://doi.org/10.3389/fpls.2025.1651814

[19] Hosoya S, Kikuchi K, Nagashima H, et al. Assessment of genetic diversity in Coho salmon (*Oncorhynchus kisutch*) populations with no family records using ddRAD-seq. BMC Res Notes, 2018, 11: 548. https://doi.org/10.1186/s13104-018-3663-4

[20] Larson WA, Seeb JE, Pascal CE, et al. Single-nucleotide polymorphisms (SNPs) identified through genotyping-by-sequencing improve genetic stock identification of Chinook salmon (*Oncorhynchus tshawytscha*) from western Alaska. Can J Fish Aquat Sci, 2014, 71: 698-708. https://doi.org/10.1139/cjfas-2013-0502

[21] Yang HN, Peng GX, Yan YC, et al. Application of DNA Molecular Marker Technology in Ecological Research. Research on Life Sciences, 2004, (S1): 96-101. https://doi.org/CNKI:SUN:SMKY.0.2004-S1-019

[22] Nei M. Molecular population genetics and evolution. North-Holland Publishing Company, 1975.

[23] Ding R, Yu D, Yang K, et al. Chromosome-Level Genome Assembly and Whole-Genome Resequencing Revealed Contrasting Population Genetic Differentiation of Black Bream (*Megalobrama skolkovii*) (Teleostei: Cyprinidae) Allopatric and Sympatric to Its Kin Species. Ecol Evol, 2025, 15(1): e70874. https://doi.org/10.1002/ece3.70874

[24] Dalvit C, Demarchi M, Cassandro M. Genetic traceability of livestock products: A review. Meat Science, 2007, 77(4): 437-49. https://doi.org/10.1016/j.meatsci.2007.05.027

[25] Nazareno AG, Bemmels JB, Dick CW, et al. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Molecular Ecology Resources, 2017, 17(6): 1136-1147. https://doi.org/10.1111/1755-0998.12654

[26] Chi BJ, Chang YM, Yan XC, et al. Analysis of Genetic Diversity and Genetic Structure between the Dalai Lake Population and the Ussuri River Population of *Leuciscus waleckii*. Journal of Fishery Sciences of China, 2010, 17(02): 228-235. https://doi.org/CNKI:SUN:ZSCK.0.2010-02-007